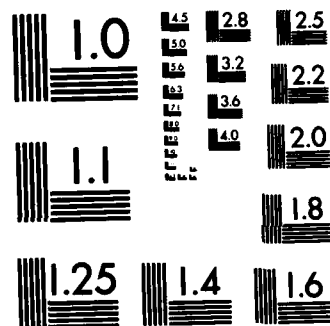END

FILMED

DTIC

MICROCOPY RESOLUTION TEST CHART

NATIONAL BUREAU OF STANDARDS-1963-A

NUSC Technical Report 6843
1 December 1982

On Generation of Random Numbers with Specified Distributions or Densities

Albert H. Nuttall
Surface Ship Sonar Department

AD A122724

DTIC
ELECTE
DEC 28 1982
S   D

D

Naval Underwater Systems Center
Newport, Rhode Island / New London, Connecticut

82 12 27 122

Preface

This research was conducted under NUSC Project No. A75205, Subproject No. ZR0000101, "Applications of Statistical Communication Theory to Acoustic Signal Processing," Principal Investigator, Dr. Albert H. Nuttall (Code 3302), Program Manager, Mr. Robert M. Hillyer, Naval Material Command (MAT 05).

The Technical Reviewer for this report was Edward M. Harris (Code 3332).

W A Von Winkle

Reviewed and Approved: 1 December 1982

W. A. Von Winkle
Associate Technical Director for Technology

The author of this report is located at the
New London Laboratory, Naval Underwater Systems Center,
New London, Connecticut 06320.

REPORT DOCUMENTATION PAGE

1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
TR 6843	AD +122 724	

4. TITLE (and Subtitle)	5. TYPE OF REPORT & PERIOD COVERED
ON GENERATION OF RANDOM NUMBERS WITH SPECIFIED DISTRIBUTIONS OR DENSITIES	
	6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s)	8. CONTRACT OR GRANT NUMBER(s)
Albert H. Nuttall	

9. PERFORMING ORGANIZATION NAME AND ADDRESS	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
Naval Underwater Systems Center New London Laboratory New London, Connecticut 06320	A75205 ZR0000101

11. CONTROLLING OFFICE NAME AND ADDRESS	12. REPORT DATE
Naval Material Command (MAT 05) Washington, D. C. 20362	1 December 1982
	13. NUMBER OF PAGES
	42

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report)
	UNCLASSIFIED
	15a. DECLASSIFICATION / DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Random Number Generation	Independent Random Variables
Probability Density	Nonlinear Distortion
Cumulative Distribution	Inverse Functions
Characteristic Function	Combinations of Random Variables
	Statistical Tests

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Generation of random numbers with specified probability density functions or cumulative distribution functions is reviewed and employed to generate some standard random variables with common densities and distributions. Combinations of random variables then afford a quick method of generating variables with more-involved distribution functions. Application to the Q_M cumulative distribution function is made as an example. Timing results for all cases are listed. Numerous statistical tests on the Hewlett-Packard 9845 Desk Calculator confirm it as a reliable generator of uniformly
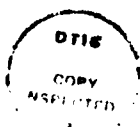
DD FORM 1473

20. (CONT'D)

distributed statistically independent random numbers.

TABLE OF CONTENTS

Accession For

NTIS GRA&I	☒
DTIC TAB	☐
Unannounced	☐
Justification	

By
Distribution/
Availability Codes

Dist	Avail and/or Special
A	

DTIC
COPY
INSPECTED

i

LIST OF TABLES

LIST OF SYMBOLS

x	random variable uniformly distributed over $(0,1)$
$g(x)$	monotonic nonlinear distortion of x
y	result of distortion g
t	nonrandom threshold
$\text{Prob}\{\}$	probability of $\{\}$
P_y	cumulative distribution function of random variable y
p_y	probability density function of random variable y
f_y	characteristic function of random variable y
$\tilde{h}$	inverse function to h
Φ	Gaussian cumulative distribution function; (18)
E	expectation; ensemble average
d	additive constant; (36)
Q_M	Q_M function; (50)
K_ν	modified Bessel function of second kind and order ν; (54)
R_k	sample correlation; (65)
μ_k	sample moment; (73)
q_k	sample probability; (77)
DFT	discrete Fourier transform

ON GENERATION OF RANDOM NUMBERS WITH SPECIFIED
DISTRIBUTIONS OR DENSITIES

INTRODUCTION

The generation of random numbers with a specified probability density function or a specified cumulative distribution function is a frequent occurrence in the simulation of signal processing techniques that are difficult or impossible to evaluate analytically. Accordingly, it is of interest to be able to generate easily and efficiently such random numbers. It is assumed that a uniform random number generator is already available; that is, independent random variables with a probability density function equal to unity over the range (0,1) can be generated. Validation of this assumption for a particular random number generator must be confirmed by numerical tests, several of which are tested herein. This report is an extension and elaboration of [1].

NONLINEAR DISTORTION OF A SINGLE UNIFORM RANDOM VARIABLE

Suppose that random variable x is uniformly distributed on (0,1). Let the nonlinear distortion of x yield the random variable y given by

$$y = g(x) = \begin{Bmatrix} g_i(x) \\ \text{or} \\ g_d(x) \end{Bmatrix} \quad , \tag{1}$$

where

$g_i(\ )$ is a monotonically increasing function of its argument,

$g_d(\ )$ is a monotonically decreasing function of its argument. $\tag{2}$

This restriction of the nonlinear distortion to be monotonic (either increasing or decreasing) is not necessary; it is done here only for simplicity. Then since random variable x is uniformly distributed on (0,1), for a fixed (nonrandom) t in the range (0,1),

$$t = \text{Prob} \left\{ x < t \right\} = \text{Prob} \begin{Bmatrix} g_i(x) < g_i(t) \\ \text{or} \\ g_d(x) > g_d(t) \end{Bmatrix}$$

$$= \text{Prob} \begin{Bmatrix} y < g_i(t) \\ \text{or} \\ y > g_d(t) \end{Bmatrix} = \begin{Bmatrix} P_y(g_i(t)) \\ \text{or} \\ 1 - P_y(g_d(t)) \end{Bmatrix} , \tag{3}$$

where we used (2), (1), and denoted the cumulative distribution function of random variable y by $P_y(\ )$. We rewrite (3) as

$$t = P_y(g_i(t))$$

$$\text{or} \tag{4}$$

$$1 - t = P_y(g_d(t)) \quad ,$$

and observe that cumulative distribution function $P_y(\ )$ is a monotonically increasing function of its argument.

Now, for arbitrary monotonic function h( ), we denote its inverse function by $\tilde{h}(\ )$; that is,*

$$\tilde{h}(h(u)) = u, \qquad h(\tilde{h}(v)) = v \quad . \tag{5}$$

Applying monotonic function $\tilde{P}_y(\ )$ to both sides of (4), and using (5), there follows

$$g_i(t) = \tilde{P}_y(t)$$

$$\text{or} \tag{6}$$

$$g_d(t) = \tilde{P}_y(1 - t) \quad .$$

Employing (6) in (1) finally yields the random variable

$$y = g(x) = \begin{cases} g_i(x) = \tilde{P}_y(\tilde{x}) \\ \qquad \text{or} \\ g_d(x) = \tilde{P}_y(1 - x) \end{cases} \tag{7}$$

as the desired result of this nonlinear transformation of random variable x. If cumulative distribution function $P_y(\ )$ of random variable y is specified and desired, and if random variable x is uniformly distributed on (0,1), (7) tells us that we must evaluate the inverse of the given cumulative distribution function and use it as the nonlinear transformation of either random variable x or random variable 1-x, depending on whether we want a monotonically increasing or monotonically decreasing transformation, respectively. The key element to this approach is the ease with which the inverse function $\tilde{P}_y(\ )$ can be computed.

---

* Some additional relationships among functions and their inverses are presented in appendix A.

EXAMPLES OF SINGLE-VARIATE DISTORTION

The following examples have been adjusted to a convenient scale, such as zero mean or unit variance. By addition of a constant and/or multiplication by a scale factor, alternative desired ranges can be realized. The times of execution given below were obtained for the Hewlett-Packard 9845B Desk Calculator equipped with the Fast Processor Upgrade Kit; they include the time required to generate the uniform random variable x. Results were obtained by averaging over 1000 independent trials. Loop counters were declared INTEGER for maximum speed.

EXPONENTIAL DENSITY

The desired probability density function of random variable y is exponential*:

$$p_y(u) = \exp(-u) \quad \text{for} \quad u > 0 \quad . \tag{8}$$

The corresponding cumulative distribution function is

$$P_y(t) = \int_{-\infty}^{t} du\, p_y(u) = 1 - \exp(-t) \quad \text{for} \quad t > 0 \quad . \tag{9}$$

The characteristic function of random variable y is (for future reference)

$$f_y(\xi) = \int_{-\infty}^{+\infty} du\, \exp(i\xi u)\, p_y(u) = (1 - i\xi)^{-1} \quad . \tag{10}$$

In order to find the inverse function to (9), we let

$$u = P_y(t) = 1 - \exp(-t) \quad . \tag{11}$$

Solving the left relation for t, and then solving the outside relation for t, we get, respectively,

---

* The probability density functions and cumulative distribution functions are zero in the unspecified regions; for example, $p_y(u) = 0$ for $u < 0$.

$$t = \tilde{P}_y(u), \quad t = -\ln(1 - u) \quad . \tag{12}$$

Combining these two, there follows for the inverse function

$$\tilde{P}_y(u) = -\ln(1-u) \quad \text{for} \quad 0 < u < 1 \quad . \tag{13}$$

The nonlinear transformation, by reference to (7), is then

$$y = \left\{ \begin{array}{c} -\ln(1-x) \\ \text{or} \\ -\ln x \end{array} \right\} \quad . \tag{14}$$

The time of execution for generation of a random variable y via (14) is 1.9 msec.

RAYLEIGH DENSITY

$$p_y(u) = u \, \exp(-u^2/2) \quad \text{for} \quad u > 0 \quad ,$$

$$P_y(u) = 1 - \exp(-u^2/2) \quad \text{for} \quad u > 0 \quad . \tag{15}$$

Via a procedure similar to (11) and (12), there follows

$$\tilde{P}_y(u) = \sqrt{-2 \, \ln(1-u)} \quad \text{for} \quad 0 < u < 1 \quad . \tag{16}$$

The nonlinear transformation that yields Rayleigh random variables is

$$y = \left\{ \begin{array}{c} \sqrt{-2 \, \ln(1-x)} \\ \text{or} \\ \sqrt{-2 \, \ln x} \end{array} \right\} \quad . \tag{17}$$

The time of execution of (17) is 2.5 msec.

5

GAUSSIAN DENSITY

$$p_y(u) = (2\pi)^{-1/2} \exp(-u^2/2) \quad \text{for all } u,$$

$$P_y(u) = \int_{-\infty}^{u} dt \, (2\pi)^{-1/2} \exp(-t^2/2) \equiv \tilde{\Phi}(u) \quad , \tag{18}$$

$$f_y(\xi) = \exp(-\xi^2/2) \quad .$$

The inverse function is [2; 26.2, 26.2.22, 26.2.23]

$$\tilde{P}_y(u) = \tilde{\tilde{\Phi}}(u) = -\tilde{\tilde{\Phi}}(1-u) \text{ for } 0 < u < 1 \quad . \tag{19}$$

The nonlinear transformation that yields Gaussian random variables from uniformly distributed ones is then, from (7) and (19),

$$y = \left\{ \begin{array}{c} \tilde{\tilde{\Phi}}(x) \\ \text{or} \\ -\tilde{\tilde{\Phi}}(x) \end{array} \right\} \quad . \tag{20}$$

The time of execution of (20) is 13.2 msec.

A much better approach, for this particular case of generation of Gaussian random variables, is as follows: let x_1 and x_2 be two independent random variables, uniformly distributed on $(0,1)$. Then according to the previous example, the two independent random variables

$$r = \sqrt{-2 \ln x_1} \quad , \quad \theta = 2\pi x_2 \quad , \tag{21}$$

have, respectively, the probability density functions

$$p_r(u) = u \exp(-u^2/2) \quad \text{for} \quad u > 0 \quad , $$

$$p_\theta(u) = (2\pi)^{-1} \quad \text{for} \quad 0 < u < 2\pi \quad . \tag{22}$$

6

Now define two new random variables by the nonlinear transformations

$$y_1 = r \cos \theta, \quad y_2 = r \sin \theta \quad . \tag{23}$$

The joint characteristic function of y_1 and y_2 is [2; 9.1.21 and 11.4.29]

$$
\begin{aligned}
E\{\exp(i\xi y_1 + i\mu y_2)\} &= E\{\exp(i\xi r \cos \theta + i\mu r \sin \theta)\} \\
&= \iint_{-\infty}^{+\infty} du\, dv\, p_r(u)\, p_\theta(v) \exp(i\xi u \cos v + i\mu u \sin v) \\
&= \int_0^{+\infty} du\, u \exp(-u^2/2) \int_0^{2\pi} dv\, \frac{1}{2\pi} \exp(iu(\xi \cos v + \mu \sin v)) \\
&= \int_0^{+\infty} du\, u \exp(-u^2/2)\, J_0\left(u\sqrt{\xi^2 + \mu^2}\right) \\
&= \exp\left(-\tfrac{1}{2}\xi^2 - \tfrac{1}{2}\mu^2\right) \quad .
\end{aligned}
\tag{24}
$$

Thus y_1 and y_2 are independent Gaussian random variables, each with zero-mean and unit variance. The time of execution of (21) and (23) is 5.4 msec per random variable (actually 10.7 msec for a pair of independent Gaussian random variables.) This 5.4 msec is considerably less than the 13.2 msec required of (20), which also requires a special function definition. A more general distortion than (21)-(23) is considered in appendix B.

Another alternative for the generation of approximately Gaussian random variables is to sum N independent random variables, uniformly distributed over (0,1). By subtraction of the constant N/2 and scaling by $\sqrt{12/N}$, a zero-mean unit-variance random variable can be generated which, however, is limited to the finite range $(-\sqrt{3N}, \sqrt{3N})$. A table of execution times and range values is given below.

Table 1. Execution Times for Sum of N Independent Random Variables

N	Execution Time (msec)	Range of Values
4	2.8	-3.46, 3.46
5	3.3	-3.87, 3.87
6	3.9	-4.24, 4.24
7	4.4	-4.58, 4.58
8	4.9	-4.90, 4.90
9	5.4	-5.20, 5.20
10	5.9	-5.48, 5.48
20	11.0	-7.75, 7.75
30	16.2	-9.49, 9.49

The characteristic function of this random variable is $[\sin(R\xi)/(R\xi)]^N$ where $R = \sqrt{3/N}$, and the normalized fourth-cumulant of the random variable generated by this summation procedure is $-1.2/N$. If the non-Gaussianness can be tolerated, this summation procedure is then a viable alternative to (21) and (23) in terms of execution time, for $N \leq 9$.

CAUCHY DENSITY

$$p_y(u) = \frac{1}{\pi} \frac{1}{1 + u^2} \quad \text{for all } u,$$

$$P_y(u) = \frac{1}{2} + \frac{1}{\pi} \arctan(u),$$

$$f_y(\xi) = \exp(-|\xi|) \quad . \tag{25}$$

This is the probability density function of the ratio of two zero-mean independent Gaussian random variables. The inverse function to the cumulative distribution function is

$$\tilde{P}_y(u) = \tan\left[\pi\left(u - \frac{1}{2}\right)\right] \quad \text{for} \quad 0 < u < 1 \quad , \tag{26}$$

and the nonlinear transformation is

8

$$y = \begin{Bmatrix} \tan\left[\pi\left(x - \frac{1}{2}\right)\right] \\ \text{or} \\ -\tan\left[\pi\left(x - \frac{1}{2}\right)\right] \end{Bmatrix}. \tag{27}$$

The execution time of (27) is 2.8 msec. Equivalently, the transformation $\tan(\pi x)$ would realize the desired probability density function, although it is not monotonic in x over (0,1).

RECTIFIED CAUCHY DENSITY

$$p_y(u) = \frac{2}{\pi} \frac{1}{1 + u^2} \qquad \text{for } u > 0 \quad,$$

$$P_y(u) = \frac{2}{\pi} \text{ arc } \tan(u) \qquad \text{for } u > 0 \quad,$$

$$\tilde{P}_y(u) = \tan\left(\frac{\pi}{2} u\right) \qquad \text{for } 0 < u < 1 \quad,$$

$$y = \begin{Bmatrix} \tan\left(\frac{\pi}{2} x\right) \\ \text{or} \\ \cot\left(\frac{\pi}{2} x\right) \end{Bmatrix}. \tag{28}$$

The execution time of (28) is 2.7 msec.

REALIZATION OF DISTRIBUTION VIA COMBINATION OF SEVERAL RANDOM VARIABLES

If several independent random variables are added, their characteristic functions multiply. So if a specified probability density function or cumulative distribution function has a characteristic function which can be broken down into easily realizable terms, this observation can be utilized to generate complicated distributions with relative ease. For example, consider the following.

CHI-SQUARED DENSITY WITH 2N DEGREES OF FREEDOM

This variate is normally thought of as being generated by summing the squares of 2N zero-mean unit-variance independent Gaussian random variables. One half of this sum has the probability density function

$$p_y(u) = \frac{u^{N-1} \exp(-u)}{(N-1)!} \quad \text{for } u > 0 \quad . \tag{29}$$

The cumulative distribution function is

$$P_y(u) = 1 - \exp(-u) \sum_{n=0}^{N-1} u^n/n! \quad \text{for } u > 0 \quad . \tag{30}$$

The inverse function, $\tilde{P}_y(\ )$, to (30) is not available in closed form for $N \geq 2$, so that recourse to (7) is not reasonable.

However, the characteristic function corresponding to (29) is

$$f_y(\xi) = (1 - i\xi)^{-N} \quad . \tag{31}$$

But this is (10) multiplied by itself N times. Then (14) reveals that we can generate y according to a sum of N independent variates:

$$y = \sum_{n=1}^{N} \left\{ -\ln x_n \right\} = -\ln \left\{ \prod_{n=1}^{N} x_n \right\} \quad . \tag{32}$$

The latter form in (32) is preferable computationally; it involves N-1 multiplies and only one logarithm, and is obviously much quicker than

10

generating 2N Gaussian random variables and summing their squares. The time of execution of (32) is given below.

Table 2. Execution Times for Chi-Squared Variate of 2N Degrees of Freedom

N	Execution Time (msec)
10	8.0
20	14.1
30	20.2

If we were to use the nonlinear distortion procedure in (20) for generating Gaussian random variables, and square and add 2N such numbers to generate a Chi-squared variate, the time of execution would be 2N (13.2) msec. For the examples in table 2, these times are 264, 528, and 792 msec respectively, which are far greater than the execution times for (32).

A closely related random variable to (32) is

$$r = \sqrt{2y} = \left[-2 \ln \left\{ \prod_{n=1}^{N} x_n \right\} \right]^{1/2} \equiv g(y) \quad . \tag{33}$$

The probability density function of random variable r is

$$p_r(u) = p_y(\bar{g}(u)) \frac{d}{du} \bar{g}(u) = p_y(u^2/2) \, u$$

$$= \frac{u^{2N-1} \exp(-u^2/2)}{2^{N-1}(N-1)!} \quad \text{for} \quad u > 0 \quad . \tag{34}$$

The cumulative distribution function of r is

$$P_r(u) = 1 - \exp(-u^2/2) \sum_{n=0}^{N-1} \frac{(u^2/2)^n}{n!} \quad \text{for} \quad u > 0 \quad . \tag{35}$$

The time of execution of (33) is given in table 3.

Table 3. Execution Times for Square Root of Chi-Squared Variate

N	Execution Time (msec)
10	8.6
20	14.7
30	20.8

Another example of this combination of variates is furnished by the following.

Q DISTRIBUTION AND RICE DENSITY

Suppose that y_1 and y_2 are two independent zero-mean unit-variance Gaussian random variables, as generated via (21) and (23). Then for a constant d, the random variable

$$z = \frac{1}{2}\left[(y_1 + d)^2 + y_2^2\right] = \frac{1}{2}\left[d^2 + 2dy_1 + y_1^2 + y_2^2\right]$$

$$= \frac{1}{2}\left[d^2 + 2dr\cos\theta + r^2\right] \tag{36}$$

has characteristic function [2; 9.1.21 and 11.4.29]

$$f_z(\xi) = E\{\exp(i\xi z)\} = E\left\{\exp\left(i\xi\left(\frac{d^2}{2} + dr\cos\theta + \frac{r^2}{2}\right)\right)\right\}$$

$$= \int_0^{+\infty} du\, u\, \exp(-u^2/2)\, \exp\left[\frac{i\xi}{2}(d^2 + u^2)\right] \int_0^{2\pi} dv\, \frac{1}{2\pi}\, \exp(i\xi d u \cos v)$$

$$= \exp(i\xi d^2/2) \int_0^{+\infty} du\, u\, \exp\left[-\frac{u^2}{2}(1-i\xi)\right] J_0(\xi d u)$$

$$= (1-i\xi)^{-1} \exp\left(\frac{d^2}{2}\frac{i\xi}{1-i\xi}\right) \quad. \tag{37}$$

12

The corresponding probability density function is

$$p_z(u) = \exp\left(- \frac{d^2}{2} - u\right) I_0\left(d\sqrt{2u}\right) \quad \text{for} \quad u > 0 \quad , \tag{38}$$

as may be confirmed by direct Fourier transformation (according to (10)) of (38); see [2; 11.4.29]. The cumulative distribution function of z is

$$P_z(u) = 1 - Q(d, \sqrt{2u}) \quad \text{for} \quad u > 0 \quad , \tag{39}$$

where Marcum's Q function is defined as

$$Q(a,b) = \int_b^{+\infty} dx \; x \; \exp\left(- \frac{a^2 + x^2}{2}\right) I_0(ax) \quad . \tag{40}$$

The time of execution of (36) and (21) is 7.2 msec; there is no need to employ (23).

A closely related random variable to (36) is

$$y = \sqrt{2z} = \left(d^2 + 2dr \cos \theta + r^2\right)^{1/2} . \tag{41}$$

The probability density function of y is the Rice density

$$p_y(u) = p_z(u^2/2) \; u$$

$$= u \exp\left(- \frac{d^2 + u^2}{2}\right) I_0(du) \quad \text{for} \quad u > 0 \quad , \tag{42}$$

and the cumulative distribution function of y is

$$P_y(u) = 1 - Q(d, u) \quad \text{for } u > 0 \quad . \tag{43}$$

The time of execution of (41) and (21) is 7.5 msec; there is no need to employ (23).

These two examples, (36) and (41), would again be analytically very difficult to realize by use of (7), because the inverse functions to cumulative distribution functions (39) and (43) are not available in closed form. Furthermore, different values of d are easily accommodated in (36) and (41), whereas the inverses to (39) and (43) would necessarily involve d as a parameter.

A more general situation is encountered for the following.

Q_M DISTRIBUTION

This example is a combination of the last two. Let, as in (36) and (21),

$$v = \frac{1}{2}\left[(y_1 + d)^2 + y_2^2\right] = \frac{1}{2}[d^2 + 2dr \cos \theta + r^2] \quad , \tag{44}$$

where

$$r = \sqrt{-2 \ln x_1} \quad , \quad \theta = 2\pi x_2 \quad . \tag{45}$$

And as in (32), let

$$w = -\ln \left\{ \prod_{m=3}^{M+1} x_m \right\} \quad . \tag{46}$$

All the random variables, $\left\{x_m\right\}_1^{M+1}$, are independent and uniformly distributed on $(0,1)$. Then let random variable z be the sum of the above two:

$$z = v + w = \frac{1}{2}[d^2 + 2dr \cos \theta + r^2] - \ln \left\{ \prod_{m=3}^{M+1} x_m \right\} \quad . \tag{47}$$

The characteristic function of random variable z follows upon use of (37) and (31):

$$f_z(\zeta) = (1-i\zeta)^{-M} \exp\left(\frac{d^2}{2}\frac{i\zeta}{1-i\zeta}\right) \quad . \tag{48}$$

The corresponding probability density function and cumulative distribution function for z are

$$p_z(u) = \left(\frac{u}{d^2/2}\right)^{\frac{M-1}{2}} \exp\left(-\frac{d^2}{2} - u\right) I_{M-1}(d\sqrt{2u}) \quad \text{for} \quad u > 0 \quad ,$$

$$P_z(u) = 1 - Q_M(d,\sqrt{2u}) \quad \text{for} \quad u > 0 \quad , \tag{49}$$

where the Q_M function is defined by [3]

$$Q_M(a,b) = \int_b^{+\infty} dx \left(\frac{x}{a}\right)^{M-1} x \exp\left(-\frac{a^2 + x^2}{2}\right) I_{M-1}(ax) \quad . \tag{50}$$

The execution time of (45) and (47) is given in table 4 for several values of M. The use of (7) would have required the inverse of (49), a rather formidable analytical task involving the two parameters d and M.

Table 4. Execution Times for Q_M Variate

M	Execution Time (msec)
10	14.6
20	20.6
30	26.7

A closely related random variable to (47) is

$$y = \sqrt{2z} = \left[d^2 + 2dr \cos\theta + r^2 - 2 \ln\left\{\prod_{m=3}^{M+1} x_m\right\}\right]^{1/2} \quad . \tag{51}$$

15

There follows

$$p_y(u) = u\left(\frac{u}{d}\right)^{M-1} \exp\left(-\frac{d^2 + u^2}{2}\right) I_{M-1}(du) \text{ for } u > 0 \quad,$$

$$P_y(u) = 1 - Q_M(d,u) \quad \text{ for } u > 0 \quad. \tag{52}$$

The execution times for (45) and (51) were 0.4 msec larger than those given in table 4.

PRODUCTS OF RANDOM VARIABLES

We can now take products of some of the random variables above and generate additional cases of complicated probability density functions. For example, if we take the product of two random variables as given by (32),

$$z = y_1 y_2 = \ln\left\{\prod_{n=1}^{N_1} x_n^{(1)}\right\} \ln\left\{\prod_{n=1}^{N_2} x_n^{(2)}\right\} \quad , \tag{53}$$

where $\left\{x_n^{(1)}\right\}$ and $\left\{x_n^{(2)}\right\}$ are uniformly distributed, the probability density function of the product random variable is, from (29) and [4; 3.471 9],

$$p_z(u) = \int_{-\infty}^{+\infty} \frac{dt}{|t|} \; p_{y_1}(t) \, p_{y_2}\left(\frac{u}{t}\right)$$

$$= \int_{0}^{+\infty} dt \; \frac{t^{N_1-N_2-1} \, u^{N_2-1}}{(N_1-1)! \; (N_2-1)!} \; \exp\left(-t - \frac{u}{t}\right)$$

$$= \frac{2 \, u^{\frac{N_1+N_2}{2}-1}}{(N_1-1)! \; (N_2-1)!} \; K_{N_1-N_2}(2\sqrt{u}) \quad \text{for} \quad u > 0 \quad . \tag{54}$$

Here $K_\nu(\;)$ is a modified Bessel function of the second kind and order ν [2; section 9.6]; it is even in ν [2; 9.6.6]. Execution times for (53) can be determined from (32) and table 2.

The random variable $w = 2z^{1/2}$, where z is given by (53), has the probability density function

$$p_w(u) = \frac{2(u/2)^{N_1+N_2-1}}{(N_1-1)! \; (N_2-1)!} \; K_{N_1-N_2}(u) \quad \text{for} \quad u > 0 \quad . \tag{55}$$

The product of two independent zero-mean Gaussian random variables, as given by (21) and (23), is given more simply by

$$z = y_1 y_2 = r^2 \sin\theta \cos\theta = \frac{1}{2} r^2 \sin(2\theta) = -\ln(x_1) \sin(4\pi x_2) \quad . \tag{56}$$

The probability density function of z can be determined exactly as in (54), with the result

$$p_z(u) = \frac{1}{\pi} K_0(|u|) \quad \text{for all } u \quad . \tag{57}$$

The 4π sweep of the sin argument in (56) is unnecessary; the following will accomplish the same probability density function:

$$z = \ln(x_1) \cos(\pi x_2) \quad , \tag{58}$$

where x_1 and x_2 are uniform over (0,1). The time of execution of (58) is 6.0 msec.

GENERATION FROM A NON UNIFORMLY DISTRIBUTED RANDOM VARIABLE

Suppose that it is possible to generate a random variable x with cumulative distribution function $P_x(\ )$, and that the cumulative distribution function of

$$y = g(x) \tag{59}$$

is desired to be $P_y(\ )$. Then for $g = g_i$, a monotonically increasing function, the cumulative distribution function of x is

$$P_x(u) = \text{Prob}\left\{x < u\right\} = \text{Prob}\left\{g_i(x) < g_i(u)\right\}$$

$$= \text{Prob}\left\{y < g_i(u)\right\} = P_y(g_i(u)) \quad . \tag{60}$$

Applying inverse function $\tilde{P}_y$ to both sides of (60), the desired nonlinearity is

$$g_i(u) = \tilde{P}_y(P_x(u)) \quad . \tag{61}$$

Equation (7) is obviously the special case of this when x is uniformly distributed. Alternative expressions to (61) are available by use of some of the relations in appendix A.

If we want to use distortion $g = g_d$, a monotonically decreasing function, to generate y, we have

$$P_x(u) = \text{Prob}\left\{x < u\right\} = \text{Prob}\left\{g_d(x) > g_d(u)\right\}$$

$$= \text{Prob}\left\{y > g_d(u)\right\} = 1 - P_y(g_d(u)) \quad . \tag{62}$$

Inverting this relation,

$$g_d(u) = \tilde{P}_y\left(1 - P_x(u)\right) \quad . \tag{63}$$

This also reduces to (7) for a uniformly distributed random variable x. In either case, (61) or (63), the inverse function of the desired cumulative distribution function, P_y, must be realized. Then a cascade operation as dictated by (61) or (63) must be employed. The desired random variable is given by

$$y = \begin{Bmatrix} \tilde{P}_y\big(P_x(x)\big) \\ or \\ \tilde{P}_y\big(1 - P_x(x)\big) \end{Bmatrix} .$$

(64)

,

TESTS OF UNIFORM RANDOM NUMBER GENERATOR

A critical component of the procedures described above is the generator of random variables, $\{x_n\}$, with a first-order probability density function that is flat over $(0,1)$, and with statistically independent samples. Here we will investigate several tests of the random number generator, RND, of the Hewlett-Packard 9845 Desk Calculator. Inherent in this investigation is the need to state the confidence associated with a particular measurement or estimate; for example, see [5].

CORRELATION TEST

The sample correlation of a set of N measurements $\{y_n\}_1^N$ is defined here as

$$R_k = \frac{1}{N-k} \sum_{n=k+1}^{N} y_n y_{n-k} \quad \text{for} \quad 0 \leq k < N \quad . \tag{65}$$

We presume that the $\{y_n\}$ are all independent with a flat probability density function

$$p_y(u) = \frac{1}{2\sqrt{3}} \quad \text{for} \quad |u| < \sqrt{3} \quad . \tag{66}$$

These random variables can be obtained from uniform $\{x_n\}$ according to the linear transformation

$$y_n = 2\sqrt{3} \left(x_n - \frac{1}{2} \right) ; \tag{67}$$

they have the convenient normalized values

$$E(y_n) = 0 \quad , \quad E(y_n^2) = 1 \quad . \tag{68}$$

More generally, from (66),

$$E\left\{ y_n^{2m+1} \right\} = 0 \quad , \quad E\left\{ y_n^{2m} \right\} = \frac{3^m}{2m+1} \quad . \tag{69}$$

21

Under these assumptions, it is readily shown, by use of (69), that R_k is unbiased, that is

$$E(R_k) = \begin{cases} 1 \text{ for } k = 0 \\ 0 \text{ otherwise} \end{cases} , \tag{70}$$

and that the standard deviation of R_k is

$$\sigma(R_k) = \begin{cases} (1.25N)^{-1/2} & \text{for} \quad k = 0 \\ (N-k)^{-1/2} & \text{for} \quad 1 \le k < N \end{cases} . \tag{71}$$

Thus we expect R_k to lie within 2 or 3 standard deviations, $\sigma(R_k)$, of $E(R_k)$ most of the time; that is, the normalized random variable

$$r_k \equiv \frac{R_k - E(R_k)}{\sigma(R_k)} \tag{72}$$

should be between ± 2 most of the time, with rare excursions to ± 3, if the $\{x_n\}$ are truly independent [5]. In table 5, the results of sample runs for N=100, 1000, and 10000 are listed. They furnish no reason for rejecting the hypothesis that $\{x_n\}$ are independent. Runs for other sets of random numbers yielded results very similar to table 5. An alternative test on the whiteness of $\{x_n\}$ is given in appendix C.

Table 5. Sample Correlation Results for (72)

Delay k	r_k for N = 100	r_k for N = 1000	r_k for N = 10000
0	1.64	0.78	0.65
1	−1.04	−0.91	0.25
2	−0.26	−0.68	−0.02
3	−0.10	1.39	−0.13
4	1.44	0.31	0.27
5	−2.31	−0.30	0.47
6	−0.49	−1.20	−0.20
7	0.83	0.81	−1.40
8	0.17	0.61	−1.47
9	−0.61	−0.77	−0.91
10	−1.43	−0.45	0.48

MOMENT TEST

The sample moment of order k for a set of N measurements $\{y_n\}$ is

$$\mu_k = \frac{1}{N} \sum_{n=1}^{N} y_n^{\,k} \quad . \tag{73}$$

The mean and variance of μ_k are

$$E(\mu_k) = E\left(y_n^{\,k}\right), \quad Var(\mu_k) = \frac{1}{N} \; Var\left(y_n^{\,k}\right) \quad , \tag{74}$$

and can be obtained from (66) and (69). The random variable

$$m_k \equiv \frac{\mu_k - E(\mu_k)}{\text{Std. Dev.}(\mu_k)} \tag{75}$$

should therefore lie mostly in the range ± 2. A sample result is given in table 6, which again confirms the flat probability density function hypothesis in (66). (Although these particular runs all resulted in positive numbers, other sample runs resulted in a preponderance of negative values for m_k.)

Table 6. Sample Moment Results for (75)

Moment k	m_k for N = 100	m_k for N = 1000	m_k for N = 10000
1	1.07	1.37	0.52
2	1.64	0.78	0.65
3	1.08	1.09	0.42
4	1.49	0.67	0.89
5	0.95	1.11	0.30
6	1.21	0.51	0.94

FIRST-ORDER DISTRIBUTION TEST

Let I_k be an interval of the line segment $(0,1)$. Then let c_n be a counting variable given by

$$c_n = \begin{cases} 1 \text{ if } x_n \in I_k \\ 0 \text{ otherwise} \end{cases} \quad \text{for } 1 \le n \le N \quad . \tag{76}$$

Then an estimate of the probability P_k, that $x_n \in I_k$, is given by the quantity

$$q_k = \frac{1}{N} \sum_{n=1}^{N} c_n \quad . \tag{77}$$

This random variable has mean and variance

$$E(q_k) = P_k, \quad Var(q_k) = P_k(1-P_k)/N \quad . \tag{78}$$

Thus the random variable

$$v_k \equiv \frac{q_k - P_k}{\left[P_k(1-P_k)/N\right]^{1/2}} \tag{79}$$

should lie most often in the region ± 2 if x_n truly does have probability P_k of lying in I_k. In table 7 is displayed the results of sample runs for the case where interval $(0,1)$ is broken into 10 equal parts; that is,

$$I_k = \left(\frac{k-1}{10}, \frac{k}{10}\right) \quad \text{for} \quad 1 \le k \le 10 \quad . \tag{80}$$

Once again, there is no reason to reject the hypothesis that the random number generator is uniformly distributed over $(0,1)$.

Table 7. Sample Probability Results for (79)

Interval k	v_k for N = 1000	v_k for N = 10000	v_k for N = 100000
1	−0.74	0.67	0.50
2	1.16	−0.57	0.67
3	−1.26	−1.13	−1.02
4	−1.37	0.03	−0.96
5	−0.32	0.90	−1.26
6	−0.32	−0.73	−0.64
7	1.48	−0.27	0.53
8	−0.53	−0.13	−0.32
9	0.32	−0.07	1.87
10	1.58	1.30	0.64

SUMMARY

Several methods of generating random variables with specified cumulative distribution functions have been presented and evaluated in terms of their time of execution and efficiency of generation. They include nonlinear distortion of a single (uniformly distributed) random variable or through combinations of simply generated random variables. The former approach requires the ability to realize the inverse function to the desired cumulative distribution function, whereas the latter approach is very fruitful if the characteristic function corresponding to the specified cumulative distribution function can be broken down into a product of simpler components. Which approach to adopt depends on the particular example and the exact way that any parameters enter into the cumulative distribution function. Of course, the inverse to a cumulative distribution function can always be evaluated numerically and used as a table look-up, in order to generate transformed random variables; however, this numerical procedure would have to be repeated if any parameter of the cumulative distribution function were desired changed.

The ability to generate uniformly distributed random variables is a key component of this procedure. Several statistical tests on the Hewlett-Packard 9845 random number generator have confirmed it to have a flat probability density function and independent samples.

These techniques are useful for generating sets of random variables of size N with a specified cumulative distribution function, and then plotting the sample cumulative distribution functions as described in [6], in order to ascertain the amount of fluctuation which is typical for that set size N and in different regions of probability. Then by comparing the amount of fluctuation of a measured data set (of unknown statistics) with typical sample cumulative distribution functions of the same size and with a known specified cumulative distribution function, decisions on acceptance or rejection of a hypothesized cumulative distribution function can be made with confidence. See [6, figures 6 and 7] for illustrations of this procedure.

APPENDIX A. SOME INVERSE FUNCTION RELATIONS

In this appendix, x and y are real (non random) variables. Let $g(x)$ be a monotonic function for x in a given range, and let

$$y = g(x) \quad . \tag{A-1}$$

The inverse relation to (A-1) is, for y in the appropriate range,

$$x = \tilde{g}(y) \quad . \tag{A-2}$$

Now suppose that we cascade nonlinearities: let

$$z = h(y) = h(g(x)) \equiv f(x) \quad , \tag{A-3}$$

to yield overall nonlinearity f. Then the outer equality in (A-3) yields

$$x = \tilde{f}(z) \quad , \tag{A-4}$$

whereas the first and third terms in (A-3) yield

$$\tilde{h}(z) = g(x), \quad \text{and} \quad \tilde{g}(\tilde{h}(z)) = x \quad . \tag{A-5}$$

Combining (A-4) and (A-5), there follows for the inverse function of the cascade, (A-3),

$$\tilde{f}(z) = \tilde{g}(\tilde{h}(z)) \quad , \tag{A-6}$$

in terms of the inverses of the individual transformations.

If we combine (A-1) and (A-2), we get

$$y = g(\tilde{g}(y)) \quad . \tag{A-7}$$

Taking a derivative with respect to y, we find

$$1 = g'(\tilde{g}(y))\, \tilde{g}'(y) \quad ;$$

$$\tilde{g}'(y) = \frac{1}{g'(\tilde{g}(y))} \quad . \tag{A-8}$$

That is, the derivative of the inverse, $\tilde{g}$, can be found in terms of the inverse function and the derivative of the original function g.

Suppose that y is given in terms of x via transformation

$$y = g(\tilde{f}(x)) \quad , \tag{A-9}$$

but the inverse function $\tilde{f}$ is impossible or very difficult to obtain from given function f. A simple way of evaluating y vs x, then, is parametrically by letting

$$t = \tilde{f}(x), \text{ to get } x = f(t),\ y = g(t) \quad . \tag{A-10}$$

Now, as t is varied, f and g can be evaluated to determine y vs x. This transformation also allows evaluation of an integral involving an inverse function:

$$\int_a^b dx\, g(\tilde{f}(x)) = \int_{t_a}^{t_b} dt\, f'(t)\, g(t) \quad , \tag{A-11}$$

where

$$t_a = \tilde{f}(a), \quad t_b = \tilde{f}(b) \quad . \tag{A-12}$$

Another use of inverse functions in integral evaluation is afforded by the example

$$I = \int_a^b dx\, g(x)\, w(x) \quad , \tag{A-13}$$

where function $w(x)$ need not be monotonic or possess an inverse. $g(x)$ is assumed monotonic in (a,b). Letting

$$y = g(x), \quad x = \tilde{g}(y), \quad dx = dy \, \tilde{g}'(y) \quad , \tag{A-14}$$

there follows, for the integral in terms of inverse function $\tilde{g}$,

$$I = \int_{g(a)}^{g(b)} dy \, y \, \tilde{g}'(y) \, w(\tilde{g}(y)) \quad . \tag{A-15}$$

Integrating by parts, using identification

$$u = y \, w(\tilde{g}(y)), \quad v = \tilde{g}(y) \quad , \tag{A-16}$$

we get the alternative form

$$I = b \, g(b) \, w(b) - a \, g(a) \, w(a) - \int_{g(a)}^{g(b)} dy \, \tilde{g}(y) \, \frac{d}{dy} \left\{ y \, w(\tilde{g}(y)) \right\} \quad . \tag{A-17}$$

The special case of $w(x) = 1$ in (A-13) and (A-17), namely

$$\int_{a}^{b} dx \, g(x) = b \, g(b) - a \, g(a) - \int_{g(a)}^{g(b)} dy \, \tilde{g}(y) \quad , \tag{A-18}$$

has the geometrical interpretation in figure A-1. Equation (A-18) is the statement that $A_1 + A_2 + A_3 =$ total area $b \, g(b)$.

As an example of the application of (A-13) and (A-15), consider

$$g(x) = \arccos\sin(x), \quad \tilde{g}(y) = \sin y \quad . \tag{A-19}$$

There follows

$$\int_{a}^{b} dx \, \arcsin(x) \, w(x) = \int_{\arcsin(a)}^{\arcsin(b)} dy \, y \, \cos(y) \, w(\sin y). \tag{A-20}$$

So if w is a polynomial, the integral on the right side of (A-20) can be
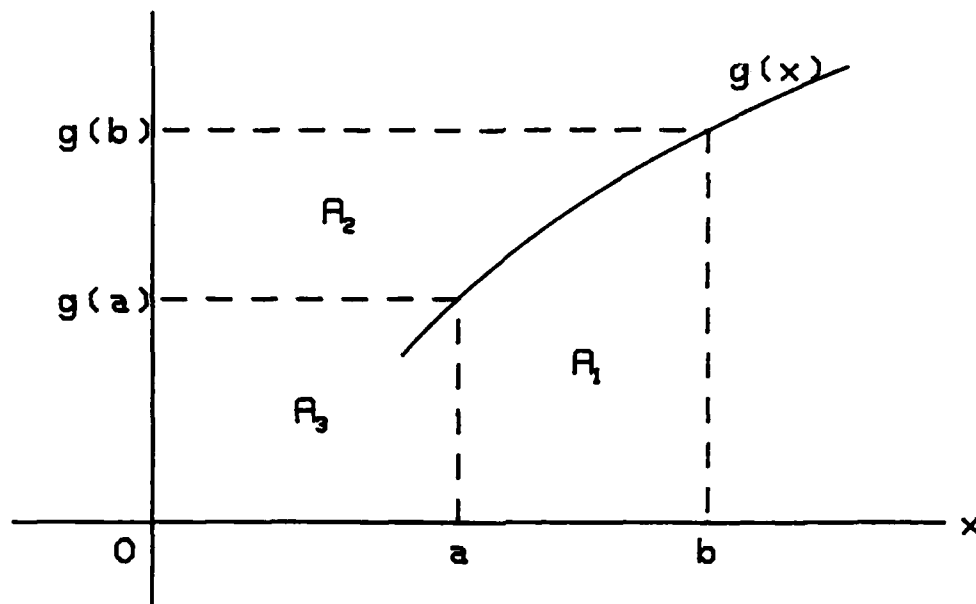evaluated in closed form.

Figure A-1. Geometrical Interpretation of (A-18)

APPENDIX. B. A MORE GENERAL DISTORTION

In (21), a random variable r with a Rayleigh probability density function and a random variable θ with a uniform probability density function were generated. These were then used in nonlinear transformation (23) to generate two independent Gaussian random variables y_1 and y_2. Here we will generalize the probability density function p_r of r in (22), and allow p_r to be arbitrary. θ is still uniform over 2π.

The two new random variables generated are again as in (23):

$$y_1 = r \cos \theta, \quad y_2 = r \sin \theta \quad . \tag{B-1}$$

Because of the uniform distribution of θ over 2π, the joint probability density function of y_1 and y_2 is of the symmetric form

$$p(y_1, y_2) = h\left(\sqrt{y_1^2 + y_2^2}\right) \quad \text{for all } y_1, y_2 \quad . \tag{B-2}$$

To determine h, observe that, for $t > 0$,

$$P \equiv \text{Prob}\left\{\sqrt{y_1^2 + y_2^2} < t\right\} = \iint_{C_t} dy_1 \, dy_2 \, p(y_1, y_2)$$

$$= \iint_{C_t} dy_1 \, dy_2 \, h\left(\sqrt{y_1^2 + y_2^2}\right) = \int_0^t d\rho \, \rho \int_{-\pi}^{\pi} d\phi \, h(\rho) = 2\pi \int_0^t d\rho \, \rho \, h(\rho) \, , \tag{B-3}$$

where C_t is a circle of radius t centered at the origin. But also, from (B-1),

$$P = \text{Prob}\left\{r < t\right\} = \int_0^t du \, p_r(u) \quad . \tag{B-4}$$

Equating (B-3) and (B-4) and taking a derivative with respect to t, there follows

$$h(t) = \frac{p_r(t)}{2\pi t} \quad \text{for } t > 0 \quad . \tag{B-5}$$

Reference to (B-2) then yields for the joint probability density function of y_1, y_2,

$$p(y_1, y_2) = \frac{p_r\left(\sqrt{y_1^2 + y_2^2}\right)}{2\pi\sqrt{y_1^2 + y_2^2}} \quad \text{for all } y_1, y_2 \quad . \tag{B-6}$$

y_1 and y_2 are statistically dependent in general.

EXAMPLE 1.

Our first case is the probability density function in (34), for random variable r as generated by (33):

$$p_r(u) = \frac{u^{2N-1} \exp(-u^2/2)}{2^{N-1}(N-1)!} \quad \text{for} \quad u > 0 \quad . \tag{B-7}$$

Substitution in (B-6) yields the joint probability density function

$$p(y_1, y_2) = \left[2\pi(N-1)!\right]^{-1} \left(\frac{y_1^2 + y_2^2}{2}\right)^{N-1} \exp\left(-\frac{y_1^2 + y_2^2}{2}\right) \quad \text{for all } y_1, y_2 \quad . \tag{B-8}$$

The special case of N=1 reduces to the pair of Gaussian random variables already considered in (21)-(24). For N=2, (B-7) and (B-8) yield

$$p_r(u) = \frac{1}{2} u^3 \exp(-u^2/2) \quad ,$$

$$p(y_1, y_2) = \frac{y_1^2 + y_2^2}{4\pi} \exp\left(-\frac{y_1^2 + y_2^2}{2}\right) \quad . \tag{B-9}$$

EXAMPLE 2.

Here random variable r is generated according to (32), and has the probability density function given by (29):

$$p_r(u) = \frac{u^{N-1} \exp(-u)}{(N-1)!} \quad \text{for} \quad u > 0 \quad . \tag{B-10}$$
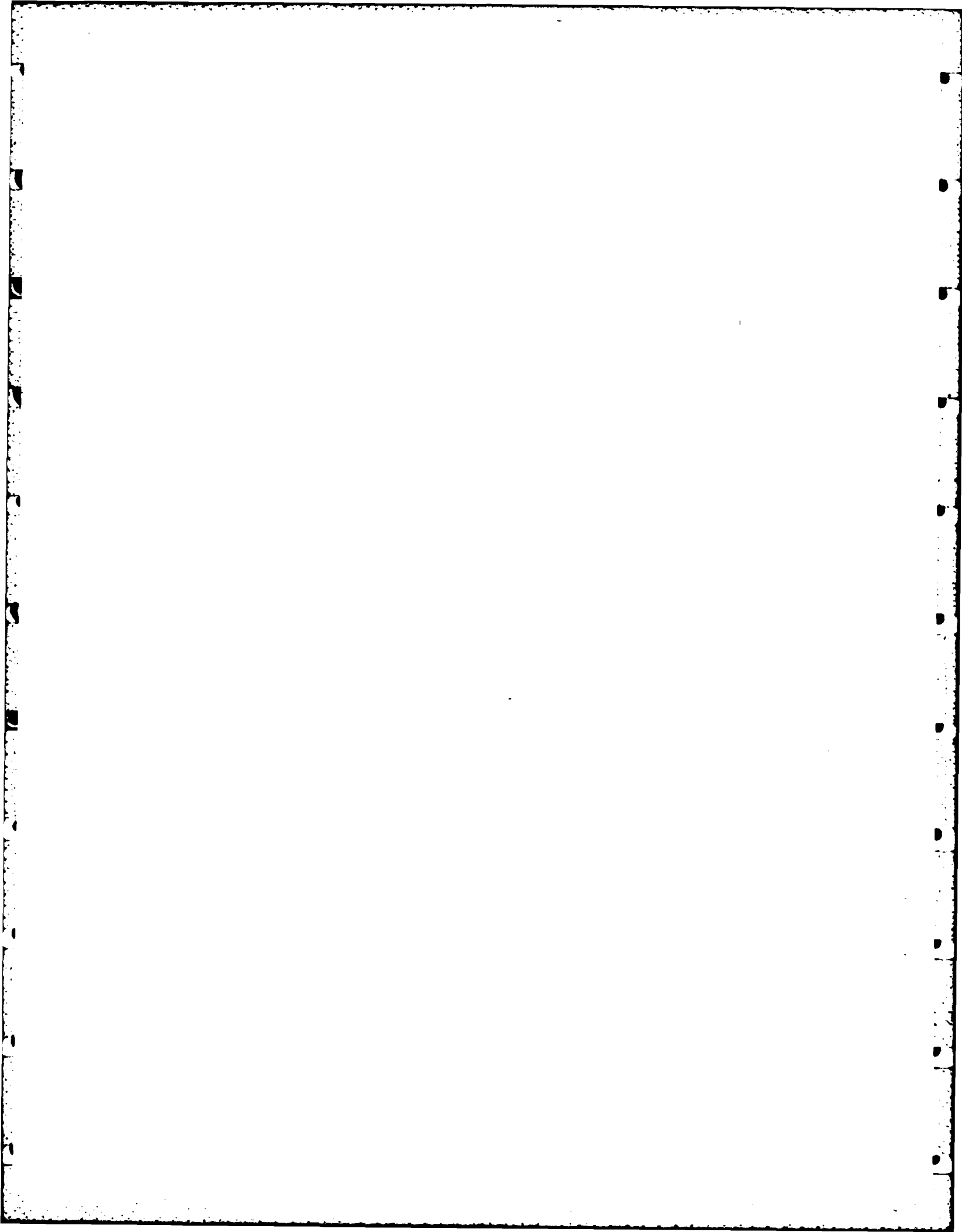
Substitution of (B-10) in (B-6) yields

$$p(y_1, y_2) = [2\pi (N-1)!]^{-1} \left(y_1^2 + y_2^2\right)^{\frac{N}{2} - 1} \exp\left(-\sqrt{y_1^2 + y_2^2}\right) \text{ for all } y_1, y_2. \quad (B-11)$$

Special cases for N equal to 1 and 2 are respectively

$$p(y_1, y_2) = (2\pi)^{-1} \left(y_1^2 + y_2^2\right)^{-1/2} \exp\left(-\sqrt{y_1^2 + y_2^2}\right) \quad (B-12)$$

and

$$p(y_1, y_2) = (2\pi)^{-1} \exp\left(-\sqrt{y_1^2 + y_2^2}\right) . \quad (B-13)$$

APPENDIX C. A TEST FOR WHITENESS OF A SEQUENCE

Suppose data points $\{x_k\}_0^{K-1}$ are available. We can form a correlation estimate at delay n according to*

$$\hat{R}_n = \frac{1}{K} \sum_k x_k\, x_{k-n}^* \quad \text{for all } n \quad , \tag{C-1}$$

where summations without limits are over the range of nonzero summands. $\hat{R}_n$ is nonzero only for $|n| < K$. If process $\{x_n\}$ were white, we would expect to have

$$\left|\hat{R}_n\right| \ll \hat{R}_0 \quad \text{for all } n \neq 0 \quad . \tag{C-2}$$

Therefore a measure of nonwhiteness is afforded by the ratio $E/\hat{R}_0^2$, where error measure E is defined as the sum of squares for $n \neq 0$,

$$E = \sum_{n \neq 0} \left|\hat{R}_n\right|^2 = \sum_n \left|\hat{R}_n\right|^2 - \hat{R}_0^2 \quad . \tag{C-3}$$

However, (C-3) is very time-consuming to calculate via (C-1), because of all the multiplications and additions required. A much more practical evaluation of (C-3) is afforded by the following procedure (the derivation is presented later in this appendix).

Define an M-point DFT of the K data points:

$$X_m = \sum_{k=0}^{K-1} x_k \exp(-i2\pi km/M) \quad \text{for} \quad 0 \leq m \leq M-1 \quad . \tag{C-4}$$

Then it follows that if $M \geq 2K-1$,

$$\sum_n \left|\hat{R}_n\right|^2 = \frac{1}{K^2 M} \sum_{m=0}^{M-1} \left|X_m\right|^4 \quad , \tag{C-5}$$

---

* The following mathematical development is very similar to that in [7].

and our whiteness measure can be expressed as

$$\frac{E}{\hat{R}_0^2} = \frac{M \sum_{m=0}^{M-1} |X_m|^4}{\left(\sum_{m=0}^{M-1} |X_m|^2\right)^2} - 1 \quad . \tag{C-6}$$

A threshold test of (C-6) is equivalent to

$$Q \equiv \frac{M \sum_{m=0}^{M-1} |X_m|^4}{\left(\sum_{m=0}^{M-1} |X_m|^2\right)^2} \quad \begin{array}{c} \text{non-white} \\ > \\ < \\ \text{white} \end{array} \quad \text{Threshold} \quad . \tag{C-7}$$

The distribution of Q could be computed from white-noise simulations, and a threshold value selected for prescribed error probability. By Schwarz's inequality, $Q \geq 1$, with equality realized if and only if $|X_m|^2$ is constant for all m.

Evaluation of Q in (C-7) requires one M-point DFT of the data $\{x_k\}_0^{K-1}$, where we must have

$$M \geq 2K - 1 \quad . \tag{C-8}$$

The subsequent calculations in (C-7) are quickly conducted.

An alternative interpretation of error measure E in (C-3) is very illuminating and lends additional credence to (C-7) as an appropriate statistic. If we define spectral estimate

$$\hat{G}_m = \sum_n \hat{R}_n \exp(-i2\pi mn/M) \quad \text{for all m} \quad , \tag{C-9}$$

then we find

$$\hat{G}_m = \frac{1}{K} |X_m|^2 \quad \text{for } 0 \leq m \leq M-1 \quad , \tag{C-10}$$

and the error measure E in (C-3) becomes

$$E = \frac{1}{M} \sum_{m=0}^{M-1} \hat{G}_m^2 - \left(\frac{1}{M} \sum_{m=0}^{M-1} \hat{G}_m \right)^2 \quad . \tag{C-11}$$

But notice that if we define the sample mean of the set of spectral estimates $\left\{ \hat{G}_m \right\}_0^{M-1}$ as

$$\mu \equiv \frac{1}{M} \sum_{m=0}^{M-1} \hat{G}_m \quad , \tag{C-12}$$

then the sample variance becomes

$$\sigma^2 \equiv \frac{1}{M} \sum_{m=0}^{M-1} (\hat{G}_m - \mu)^2$$

$$= \frac{1}{M} \sum_{m=0}^{M-1} \hat{G}_m^2 - \left(\frac{1}{M} \sum_{m=0}^{M-1} \hat{G}_m \right)^2 = E \quad . \tag{C-13}$$

That is, error measure E is equal to the sample variance of the set of spectral estimates. This latter quantity is a very natural measure for deducing whether a sequence is white, since σ^2 would be expected to be smaller for a truly white process.

This also suggests that a meaningful measure, for determining whether a sequence is white over a limited band of the zero-to-Nyquist range, is the sample variance of the spectral estimates over that particular band in question.

For a real sequence $\left\{ x_k \right\}_0^{K-1}$, the sums in (C-7) need only be conducted over half of the range, by virtue of the conjugate symmetry:

$$X_{M-m} = X_m^* \quad \text{for} \quad 1 \leq m \leq M-1 \quad \text{if} \quad \left\{ x_k \right\} \text{ real} \quad . \tag{C-14}$$

37

Thus then (for M even)

$$\sum_{m=0}^{M-1} |X_m|^4 = X_0^4 + X_{\frac{M}{2}}^4 + 2 \sum_{m=1}^{\frac{M}{2}-1} |X_m|^4 \quad \text{if} \quad \{x_k\} \text{ real.} \qquad (C-15)$$

DERIVATIONS

Define spectral estimate

$$\hat{G}_m^{(M)} = \sum_n \hat{R}_n \exp(-i2\pi nm/M) \quad \text{for all } m \quad . \qquad (C-16)$$

The superscript specifically indicates the period in the subscript variable. Substituting (C-1) in (C-16), there follows

$$\hat{G}_m^{(M)} = \sum_n \exp(-i2\pi nm/M) \frac{1}{K} \sum_k x_k \, x_{k-n}^*$$

$$= \frac{1}{K} \sum_k x_k \exp(-i2\pi km/M) \left[\sum_n x_{k-n} \exp(-i2\pi(k-n)m/M) \right]^*$$

$$= \frac{1}{K} \left| X_m^{(M)} \right|^2 \quad , \qquad (C-17)$$

where

$$X_m^{(M)} = \sum_k x_k \exp(-i2\pi km/M) \quad \text{for all } m \quad . \qquad (C-18)$$

Now the inverse DFT of $\left\{ \hat{G}_m^{(M)} \right\}$ is

$$\hat{R}_n^{(M)} \equiv \frac{1}{M} \sum_{m=0}^{M-1} \hat{G}_m^{(M)} \exp(i2\pi mn/M) = \quad \ldots + \hat{R}_{n-M} + \hat{R}_n + \hat{R}_{n+M} + \ldots \quad (C-19)$$

for all n,

which is an infinitely-aliased version of $\left\{ \hat{R}_n \right\}$. However, if

$$M \geq 2K-1, \qquad (C-20)$$

38

there is no overlap of terms in (C-19). Henceforth we presume (C-20) to be true; that is, the size of the transform (C-18) must be at least twice as large as the number of data points. Then there follows from (C-19),

$$\hat{R}_n = \frac{1}{M} \sum_{m=0}^{M-1} \hat{G}_m^{(M)} \exp(i2\pi nm/M) \quad \text{for } |n| < K \quad . \tag{C-21}$$

Therefore

$$\sum_n \left|\hat{R}_n\right|^2 = \sum_{n=-K+1}^{K-1} \left|\hat{R}_n\right|^2 = \sum_{\substack{n:\text{one period} \\ \text{of length } M}} \left|\hat{R}_n^{(M)}\right|^2$$

$$= \sum_{n=0}^{M-1} \frac{1}{M^2} \sum_{m=0}^{M-1} \sum_{p=0}^{M-1} \hat{G}_m^{(M)} \hat{G}_p^{(M)} \exp(i2\pi(m-p)n/M) = \frac{1}{M} \sum_{m=0}^{M-1} \hat{G}_m^{(M)2} \quad . \tag{C-22}$$

Then from (C-3) and (C-19),

$$E = \frac{1}{M} \sum_{m=0}^{M-1} \hat{G}_m^{(M)2} - \left(\frac{1}{M} \sum_{m=0}^{M-1} \hat{G}_m^{(M)}\right)^2 \quad . \tag{C-23}$$

Although (C-18) is defined for all relative sizes of M and K, the relation (C-23) holds only if (C-20) holds.

REFERENCES

1. A. H. Nuttall, "Some Techniques for Generation of Random Numbers with Specified Distributions or Densities," NUSC Technical Memorandum 811135, 1 October 1981.

2. Handbook of Mathematical Functions, U. S. Department of Commerce, National Bureau of Standards, Applied Mathematics Series No. 55, U. S. Government Printing Office, June 1964.

3. A. H. Nuttall, "Some Integrals Involving the Q_M-Function," NUSC Technical Report 4755, 15 May 1974.

4. I. S. Gradshteyn and I. M. Ryzhik, Table of Integrals, Series, and Products, Academic Press, Inc., N. Y., 1980.

5. K. D. Senne, "Machine Independent Monte Carlo Evaluation of the Performance of Dynamic Stochastic Systems," Journal of Stochastics, Vol. 1, pp. 215-238, September 1974.

6. A. H. Nuttall, "On Characterization of the Probability Distribution of Measured Data," NUSC Technical Memorandum 811184, 10 December 1981.

7. A. H. Nuttall, "Reconstruction of Power Spectral Estimates," NUSC Technical Memorandum 781051, 6 March 1978.

INITIAL DISTRIBUTION LIST

Addressee	No. of Copies
ASN (RE&S)	1
OUSDR&E (Research & Advanced Technology)	2
Deputy USDR&E (Res & Adv Tech)	1
OASN, Dep Assist Secretary (Res & Adv Tech)	1
ONR, ONR-100, -102, -200, -400, -410, -422, -425, -425AC, -430,	9
CNO, OP-090, -098, -902, -941, -951, -951D, -951E, -952, -96	
-981F, -981, -982F, -983	13
CNM, MAT-00, -05, SP-20, -24, ASW-12, -11, -13, -14, MAT-03621,	
Systems Project Office (PM-2) Trident,	
Surface Ships Project Office (PM-18)	11
DIA, DT-2C	1
NRL	1
NRL, USRD	1
NRL, AESD	1
NORDA	1
USOC, Code 241, 240	2
SUBBASE LANT	1
OCEANAV	1
NEOC	1
NWOC	1
NAVOCEANO, Code 02, 3400, 6200	3
NAVELECSYSCOM, ELEX 00, 03, 320	3
NAVSEASYSCOM, Sea-00, -05R, -611, -613, -62R, -63R, -63R-1,	
-631Y, -631Y (Surface Sys Sub-Gr), SEA-92R	10
NAVAIRDEVCEN	1
NAVAIRDEVCEN, Key West	1
NOSC	1
NOSC, Code 6565 (Library	1
NAVWPNSCEN	1
NCSC	1
NAVCIVENGRLAB	1
NAVSWC	1
NAV SURFACE WEAPONS CENTER, WHITE OAK LAB	1
NAVTRAEQUIPCENT Technical Library	1
NAVPGSCOL	1
NAVAIRTESTCEN	1
APL/UW, SEATTLE	1
ARL/PENN STATE, STATE COLLEGE	1
CENTER FOR NAVAL ANALYSES (ACQUISITION UNIT)	1
DTIC	1
DARPA	1
NOAA/ERL	1
NATIONAL RESEARCH COUNCIL	1
WEAPON SYSTEM EVALUATION GROUP	1
WOODS HOLE OCEANOGRAPHIC INSTITUTION	1
ENGINEERING SOCIETIES LIB. UNITED ENGRG CTR	1
ARL, UNIV OF TEXAS	1
MARINE PHYSICAL LAB, SCRIPPS	1
INTERMETRICS INC. EAST LYME, CT	1
MAR, INC. EAST LYME, CT	1
EG&G WASHINGTON ANAL SER CTR, INC. (T. Russell)	1
HYDROTRONICS INC. (D. Clark)	1
ANALYSIS & TECHNOLOGY, INC., N. STONINGTON (T. Dziedzic)	1
EDO CORPORATION, NY (J. Vincenzo)	1

END

FILMED

2-83

DTIC